# Statistical Analysis Report

## Does Work Location Affect the Isolation Score AND Stress Level?

**11/12/2025**

# TEAM MEMBERS

| Name | Role |
|---|---|
| Abanob Adly Gad | Leader |
| Anthony George Shaker | Member |
| Kirolos Samy Tawadrous | Member |
| Pierre Bassem Salah | Member |
| Sabry Ehab Wahba | Member |

# Project Overview & Descriptive Statistics & Data Exploration

## 1. Project Overview:

This project analyzes a dataset of professional employees to investigate the relationships between Work Location (Remote, Hybrid, Office), Social Isolation, Stress Levels, and Psychological Wellbeing. The analysis utilizes descriptive statistics, regression modeling, probability theory (Bayes' Theorem), and sampling distributions to derive meaningful insights.

## 2. Descriptive Statistics & Data Exploration

### 2.1 Isolation Score Analysis

- **Overall Mean Isolation Score**: **4.95**
- **Overall Median Isolation Score**: **4.72**

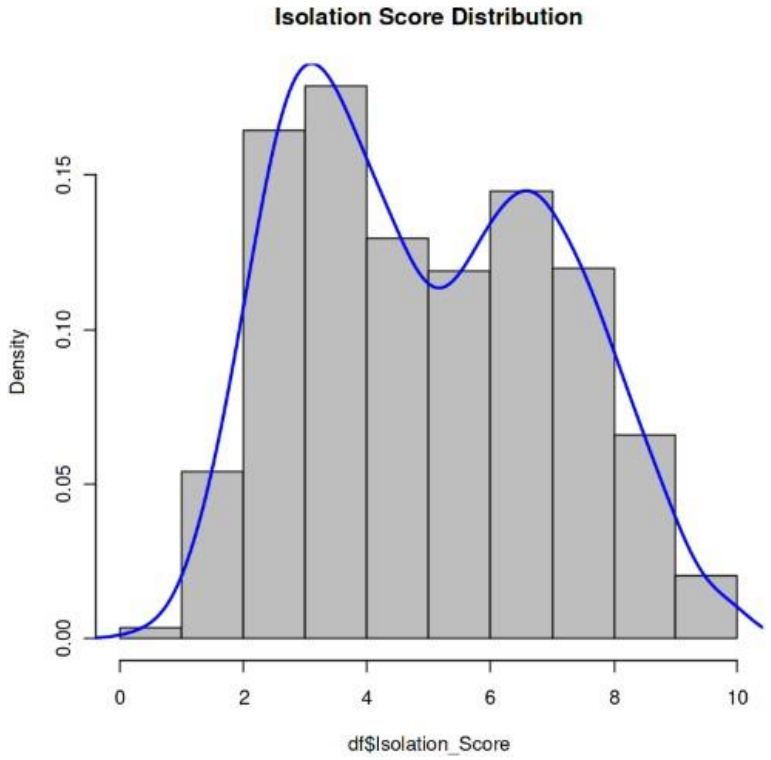



**Figure (1): The Distribution is positively Skewed to the Right(Mode<Midian<Mean) .**

## 2.2 Impact of Work Location on Isolation:

**Remote Workers**: **6.83**. (Highest average isolation)

**Hybrid Workers**: **4.17**. (Moderate average isolation)

**Office Workers**: **2.82**. (Lowest average isolation)

## 2.3 A boxplot was utilized to visualize these differences statistically.

**Observation:** The "Remote" category shows a significantly higher interquartile range (IQR) positioned much higher on the Y-axis compared to "**Office**" and "**Hybrid**."

Box Plot:

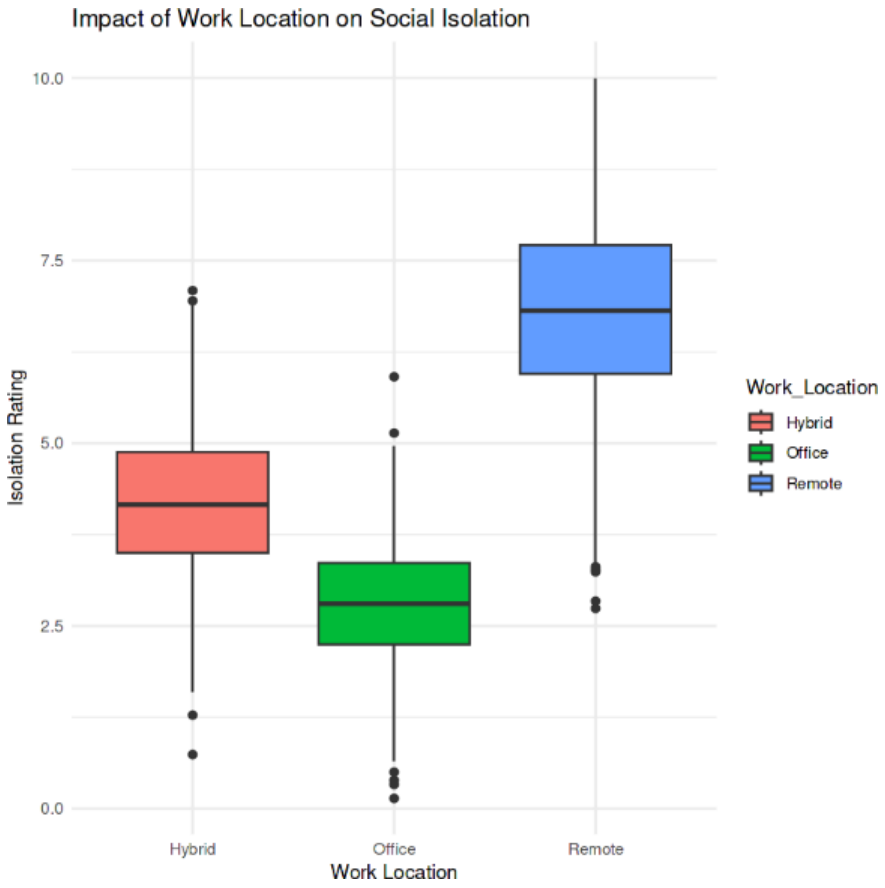



**Figure (2): There is a clear trend were working remotely significantly increases the feeling of social isolation compared to working from the office.**

# Descriptive Statistics and Inferential Statistics

## 3.0 Probability &Inferential Statistics:

**The code acts as a flexible engine**: The program inputs the parameters (like sample size or desired thresholds), and it outputs the mathematical probability of that event occurring.

## 3.1 Predicting Specific Employee Outcomes (Z-Score Methodology)

A method to determine the likelihood of any single employee falling within a specific range of isolation.

- **The Logi**c: By modelling the population's distribution using the Mean and Standard Deviation, the code creates a "Normal Distribution Curve."
- **Application:** "What is the chance a random employee scores below X?"
- **Example:** When the program tested for a score of 6.5, the model returned a probability of ~40%

## 3.2 Probability of a Single Proportion

- **Purpose:** This code is used to calculate the probability that a certain proportion of people in a sample has a specific characteristic, based on the known proportion in the entire population.
- **Result:** The output gives a probability value, which tells you how likely or unlikely it is to see the sample proportion you are evaluating.
- **When you want to check if a sample reflects the population accurately.**

## 3.3 Assessing Group Differences: Sampling Distribution of Means

The program compares the average score of two groups.

- The goal is to understand how likely it is that a random sample would show a difference that is as large as the one we are interested in (for example, a difference of 4 points or more).
    - If this probability is high, it means the difference is expected and probably reflects a real, consistent gap between the two groups.
    - If the probability is low, then such a big difference is unlikely to appear just from normal sampling, meaning the observed gap might be due to chance, not a real effect.
- Assumptions:

The samples **must be random, independent**, **and large enough** for the comparison to be reliable.

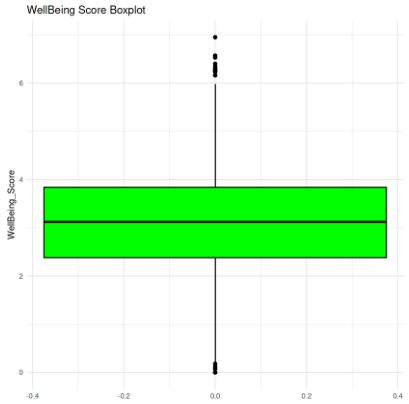
## 3.4 Difference in Proportions

- We compare the **percentage of people** with a **certain outcome** in two groups.
- The goal is to check the probability that a **random sample** would show a difference in **percentages** that is larger than a certain cutoff (like more than 5%).
    - If this probability is **high**, it means the difference between the groups is **real** and **meaningful**.
    - If the probability is **low**, then the difference **might not be stable** — it could simply be random variation rather than a true pattern.
- Assumptions: The groups **must be sampled properly**, and the **sample sizes need to be large enough** so the percentages we calculate are trustworthy.

# 4.0 Visualization
# Wellbeing Score & Social Interaction

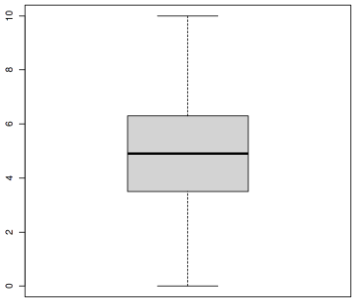
## 4.1 Wellbeing Score Analysis

- Minimum: 0.19
- First Quartile: 2.38
- Mean: 3.117874
- Median: 3.12
- Third Quartile: 3.84
- Max: 5.98
- IQR: 1.46




## 4.2 Social Interactions Analysis:

- Minimum: 0
- First Quartile: 3.5
- Mean: 4.92
- Median: 4.9
- Third Quartile: 6.3
- Max: 10
- IQR: 2.8



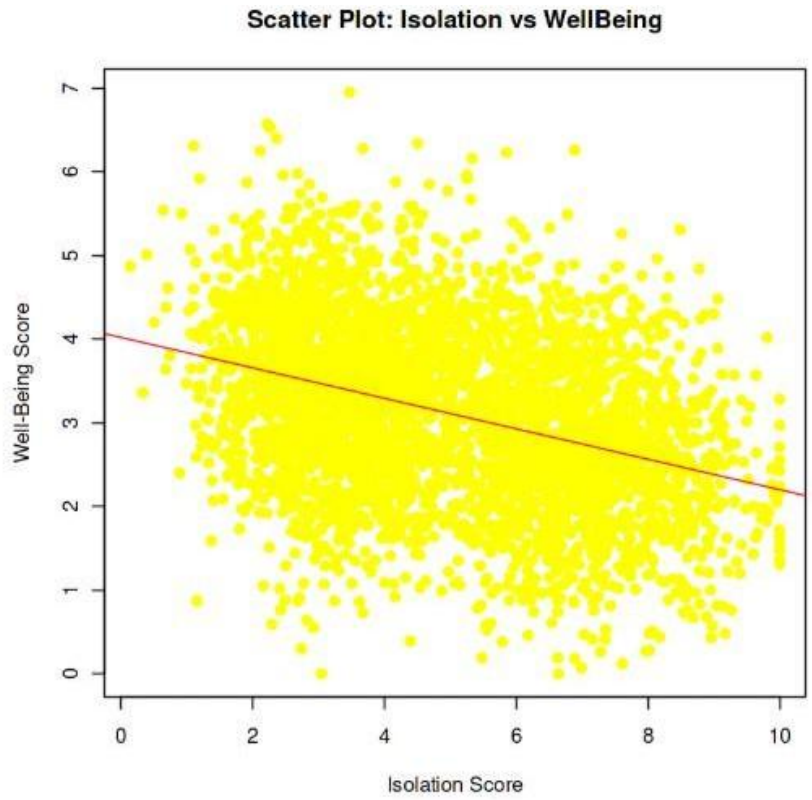
## 5.0 Correlation & Regression Analysis

### 5 .1 Relationship: Isolation vs. Wellbeing

**Pearson Correlation Coefficient (r): -0.36**

**Interpretation**: A moderate **negative correlation**.
As isolation scores go up, wellbeing scores tend to go down.

**Coefficient of Determination ($R^2$): 0.132**
**Interpretation:** Only **13.2%** of the variation in Wellbeing is explained by Isolation. This suggests that Wellbeing is a complex metric influenced by many other unmeasured factors (beyond just isolation).

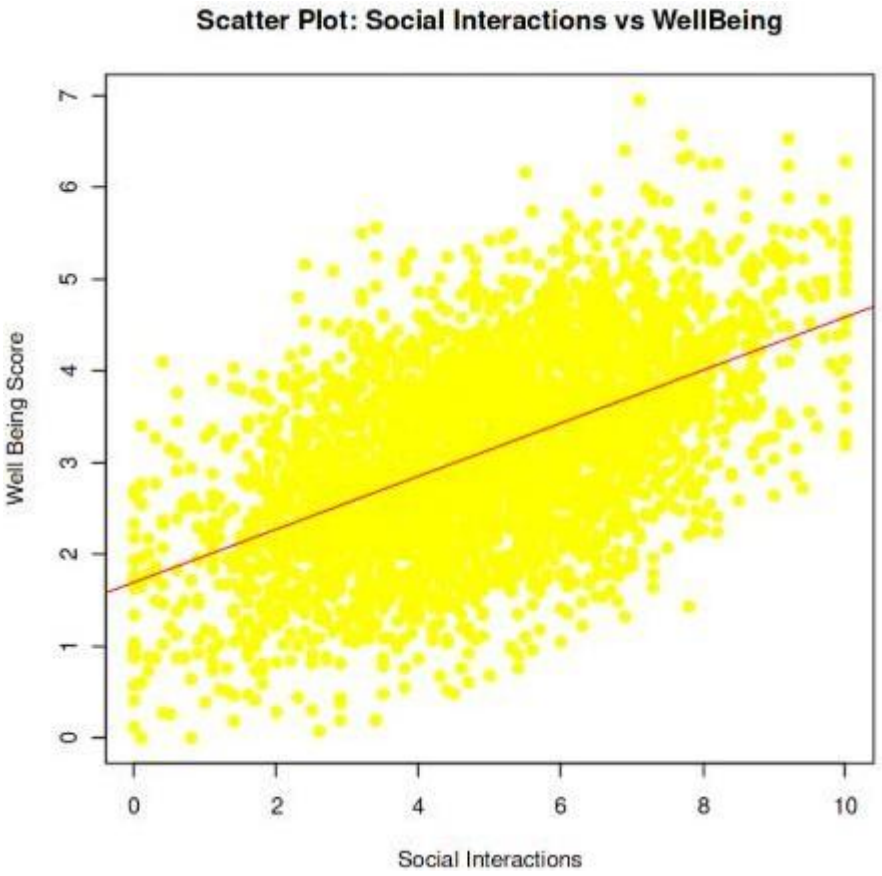


## 5.2 Relationship: Social Interaction Vs Wellbeing

Pearson Correlation Coefficient (r): 0.54
Interpretation: A moderate positive Correlation.
As social interactions go up, wellbeing scores also go up.

Coefficient of Determination ($R^2$): 0.29
Interpretation: approximately 30% of the variation in Wellbeing is explained by social interactions.

**$Y = b_0 + b_1X$**
**X=10 (Social interaction Score)**
**Y= 4.59 (Wellbeing Score)**



## 6.0 Work Location AND Stress Level

### 6.1 Contingency Table:

| | High | Medium | Total |
|---|---|---|---|
| Hybrid | 639 | 74 | 713 |
| Office | 983 | 181 | 1164 |
| Remote | 1558 | 65 | 1623 |
| Total | 3180 | 320 | 3500 |

---

### 6.2 Conditional Probability

**Purpose:** This code calculates the conditional probability of one event happening given another.

**The output** is the probability that an employee has specific stress level if we know their work location.

**Formula**:

**P(Stress/Location) = P(Stress Ω Location) / P(Location)**

## Probability Analysis & *Final Conclusions and Summary*

### 6.3 Bayes' Theorem

**Purpose**: This code uses Bayes' Rule to calculate the probability that an employee works in a certain location (e.g., Remote) given that they have a specific characteristic (e.g., high stress level).

**EXAMPLE:**
What is the probability that an employee is working Remotely, given that they are suffering from High Stress?

**P(Remote): 0.4637** (46.37% of employees are remote).

**P (High Stress): 0.9086** (90.86% of all employees report high stress.

**P (Stress |Remote): 0.9600 (**96% of remote workers are stressed).
**Result:**
**P (Remote | Stress): 0.4899**

**Conclusion**: If we randomly select an employee with High Stress, there is a 48.99% probability that this person works Remotely.

---

## *7.0 Final Conclusions & Summary*

1. The Remote Work Trade-off: While remote work offers flexibility, it is strongly associated with higher social isolation (Avg: 6.83) compared to office work (Avg: 2.82).

2. Wellbeing at Risk: **Higher isolation correlates with lower wellbeing**, though it is not the only factor involved.

3. Stress Level: High stress is very common across the dataset (90%+), but it is statistically more prevalent among remote workers.

4. Recommendation: Organizations with remote workforces should implement specific interventions to reduce isolation and manage stress to improve overall employee wellbeing.